STATS 4CI3/6CI3 Winter 2021

ASSIGNMENT 4

Submit to Crowdmark using the link that was emailed to you.

Due before 11 PM on Friday, March 12th.

Your assignment must conform to the Assignment Standards listed below.

Assignments submitted up to 24 hours late will incur a 30% penalty.
Assignments submitted more than 24 hours late will receive a zero grade.

Answer all questions, stating your answer and showing your code. Not all questions carry equal marks.

Set a seed where appropriate to make your work reproducible

1. **(15 MARKS)**
   U_1, U_2 and U_3 are $U[0,1]$ random variables. **Set the seed** to (1234) where appropriate and estimate the following expectations, using a simulation size of 1000;

   (a) $E[1/(1+U_1)]$

   (b) $E[(U_1+U_2+U_3)^{1/2}]$

2. **(22 MARKS)**
   The Central Limit Theorem states that if $X \sim Bin(m,p)$ then

   $$Z = \frac{X - mp}{\sqrt{mp(1-p)}}$$

   is approximately standard normal, under certain conditions.

   For $p = 0.4$ and $m = 10$;

   (a) Simulate (using a simulation size of 1000) a random sample of Z and state whether it looks normally distributed [hint: use the qqnorm() function]. Show the Normal Q-Q plot as part of your answer.

(b) Repeat this for $m = 100, m = 1,000$, and $m = 10,000$. When does the approximation look acceptable?

(c) What happens if you set $m = 1,000$ and $p = 0.01$?

3. **(40 MARKS)** It is possible demonstrate the existence of the stationary distribution of a Markov chain by running a simulation experiment. You start a random walk at a particular state, for example location 3, and then simulate many steps of the Markov chain using the transition matrix P. The relative frequencies of the walker in the six locations after many steps will eventually approach the stationary distribution w.

(a) Start your simulation by reading in the transition matrix P below,

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.50 0.50 0.00 0.00 0.00 0.00
[2,] 0.25 0.50 0.25 0.00 0.00 0.00
[3,] 0.00 0.25 0.50 0.25 0.00 0.00
[4,] 0.00 0.00 0.25 0.50 0.25 0.00
[5,] 0.00 0.00 0.00 0.25 0.50 0.25
[6,] 0.00 0.00 0.00 0.00 0.50 0.50
```

(b) Set up a storage vector s for the locations of the walker in the random walk. I recommend defining s as an "array". Note: you will be simulating up to 50,000 steps in the Markov chain later.

(c) Set the starting location for the walker to state 3. Write code that can simulate 50,000 draws from the Markov chain. (Use the "sample" function to simulate one step).

Hint: you should sample a single value from the set $\{1, 2, 3, 4, 5, 6\}$ with probabilities given by the $s(j-1)$ row of the transition matrix P, where $s(j-1)$ is the current location of the walker.

(d) Summarize the relative frequencies of visits to the six states after 500, 2000, 8000, and 50,000 steps of the chain using the "table" command. (Convert the counts to relative frequencies by dividing by the number of steps.)

(e) What do you think is the stationary distribution, w, of this chain? (correct to 1 decimal place)

4. **(18 MARKS)**

   (a) Produce pseudocode for a Metropolis-Hastings algorithm.

   (b) What is the difference between a Metropolis-Hastings algorithm and an Independence Metropolis-Hastings algorithm?

   (c) For an Independence Metropolis-Hastings algorithm to work, what condition must the candidate density meet?

   (d) Explain the difference in the nature of the outputs obtained from a Metropolis-Hastings algorithm as compared to the outputs obtained from an Accept-Reject algorithm. (i.e. what type of sample are you obtaining for each approach?)

   (e) What is the main consequence of selecting an inappropriate candidate density?

   (f) Compared to the chain from an Independence Metropolis-Hastings algorithm, how would you describe how the chain from a Random Walk Metropolis-Hastings algorithm moves?

   (g) Explain the importance of scale (of the candidate density) for a Random Walk Metropolis-Hastings algorithm.

   (h) What is the main difference between a Markov Chain produced with the Gibbs Sampling technique, as compared to a Markov Chain obtained using a standard Metropolis-Hastings algorithm?

   (i) What is the easiest way to avoid "false convergence" of a chain?

5. **(5 MARKS)** With regard to the LearnBayes package in R, define (i.e. provide a "Description") of the following functions:

   (a) cauchyerrorpost

   (b) gibbs

   (c) groupeddatapost

   (d) indepmetrop

   (e) rwmetrop

Assignment Standards

- LaTeX is strongly recommended but not strictly required. The use of Markdown in R studio is also recommended.
- Submit your assignment as one **.pdf document**. **All R code should be included inline.**
- Do not include a title page. The title, your **name and student number** should be printed at the top of the first page.
- The writing and referencing should be appropriate to the university level.
- Various tools, including publicly available internet tools, may be used by the instructor to check the originality of submitted work.